Road transport data and their uses

Richard Gibbens

Computer Laboratory
University of Cambridge

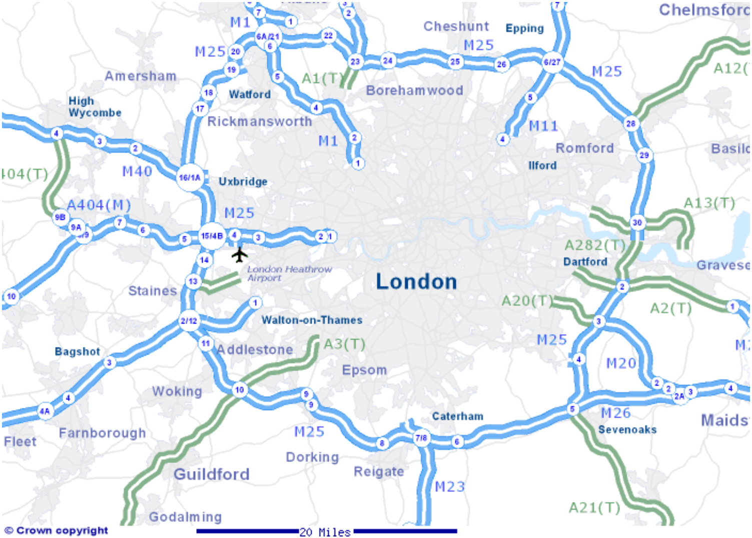
Cambridge Statistics Discussion Group
5 February 2008

MIDAS

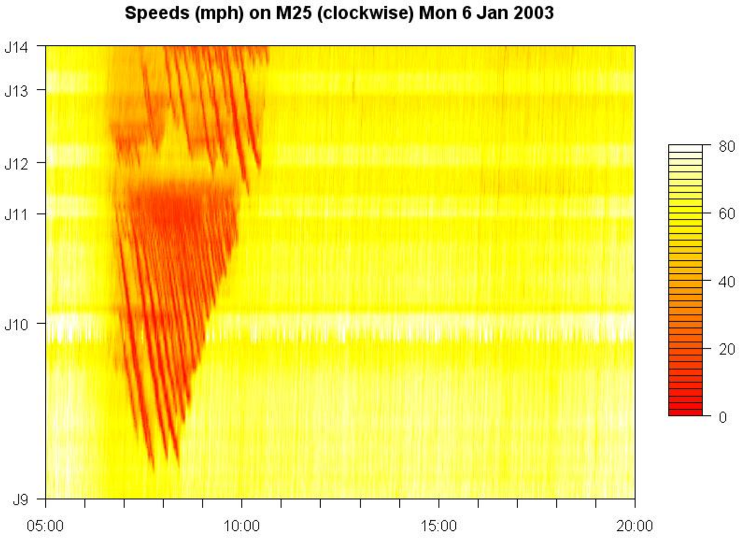
Motorway incident detection and automatic signalling

Designed for real-time closed loop control of speed limits. Presently, covers about 30% of the Highways Agency's strategic road network. Earliest data recorded in 1995.

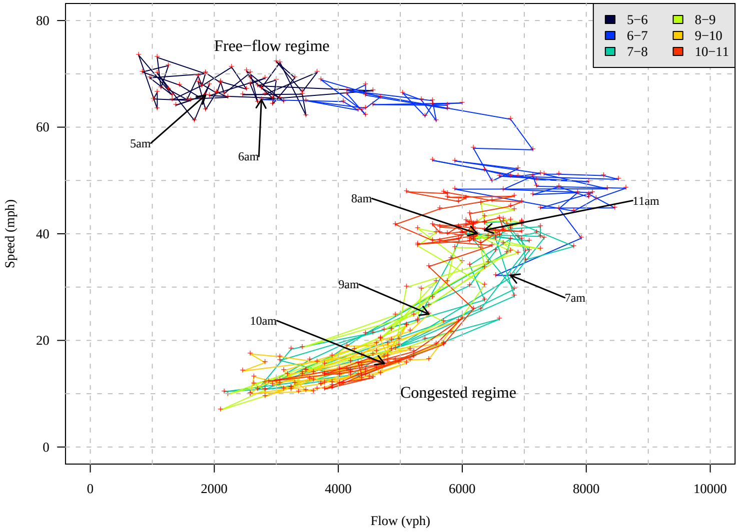
M25 motorway



M25 speeds


Speeds (mph) on M25 (clockwise) Mon 6 Jan 2003
J14
J13
J12
J11
J10
J9
05:00
10:00
15:00
20:00
80
60
40
20
0

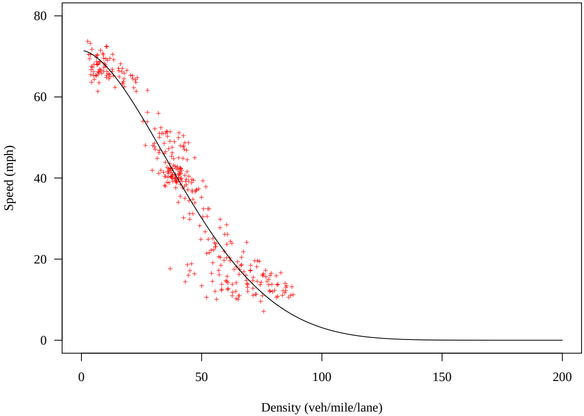
Speed/flow relationship and flow breakdown



Speed/density relationship

Consider the flow density, ρ, given by

flow (veh/hr) $=$ density (veh/mile/lane)$\times$speed (mile/hr)$\times$lanes (n) .



Nonlinear model for speed/density relationship

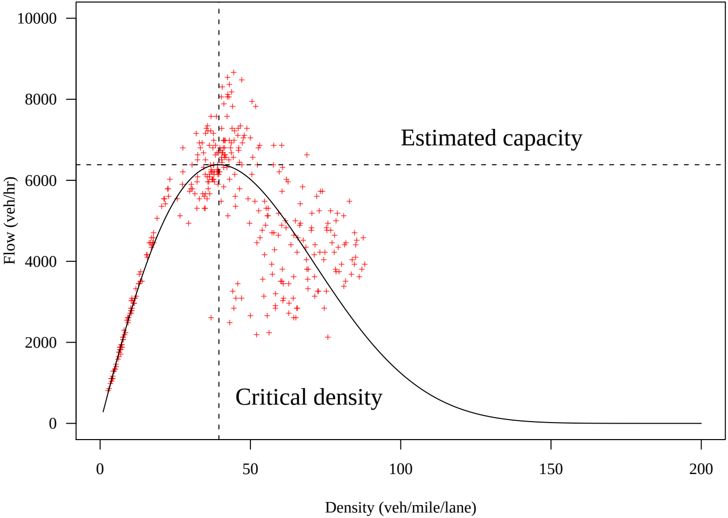
Here the speed, $s(t)$, and flow density $\rho(t)$ have been modelled by the following relationship

$$s(t) \sim s_{\text{free}} \left[1 - \left(\frac{\rho(t)}{\rho_{\text{jam}}} \right)^a \right]^b .$$

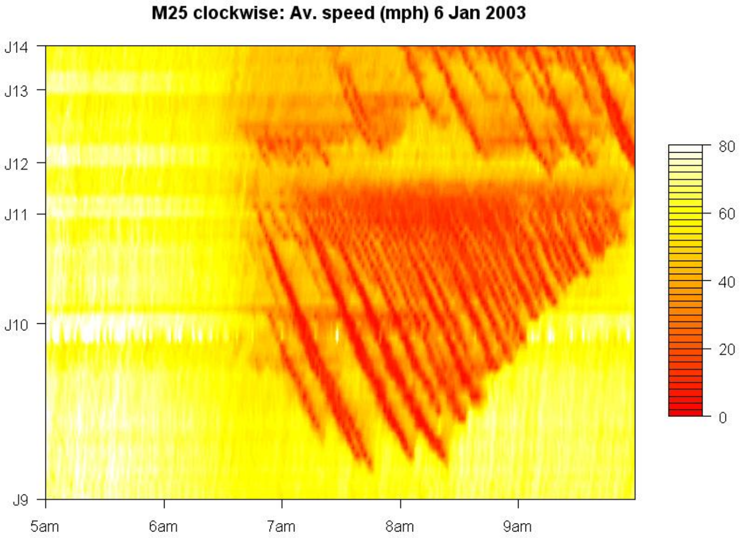
The parameters s_{free} and ρ_{jam} are the speed of free flowing vehicles and the flow density when flow eventually ceases, respectively. The parameters a and b are fitted empirically by non-linear least squares.
Background discussion can be found in Tom Bellemans' PhD thesis (2003).

Estimating capacity

Critical density ≈ 39.5 veh/mile/lane; capacity ≈ 6,385 veh/hr



M25 speeds


M25 clockwise: Av. speed (mph) 6 Jan 2003
J14
J13
J12
J11
J10
J9
5am
6am
7am
8am
9am
80
60
40
20
0

Journey times

- Journey planning
  - Generalized cost of travel
  - Traffic assignment and Wardrop equilibrium
  - Longer term planning: home and work
- Policy
  - Social cost calculations ...

Journey planning

- Generalized cost of travel $=$

  monetary costs (*fares, fuel, wear & tear, tolls*) $+$
  non-monetary costs (*value of time* $\times$ *journey time*)
- Traffic assignment and Wardrop equilibrium
  - Wardrop's (first) principle states that

    "The journey times in all routes actually used are equal
    and less than those which would be experienced
    by a single vehicle on any unused route"
- Longer term planning: journey times between home and work and land use patterns more generally

Policy

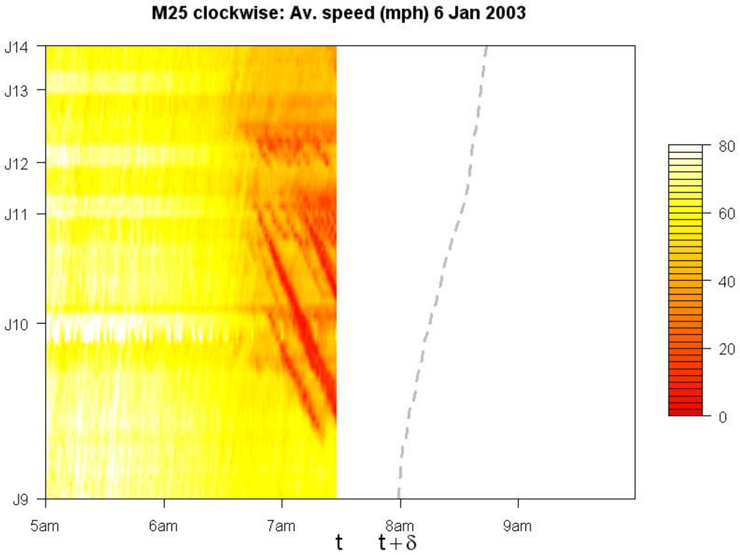
Social costs include:

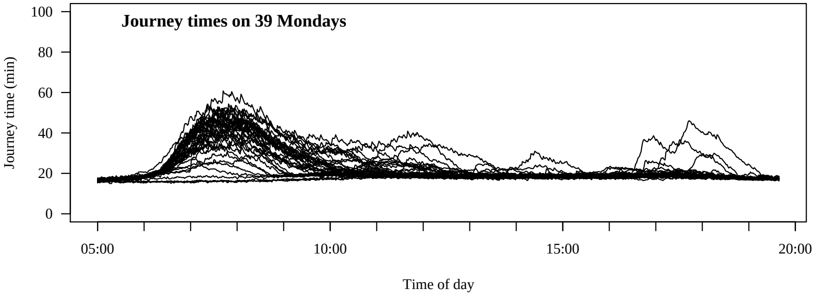
- Congestion
  - The marginal cost of congestion is the additional cost of delay incurred by all travellers when a single vehicle decides to travel a given unit distance
  - Highly variable with location and time of day
- Accidents
- Air and noise pollution
- Climate change (social cost of carbon)

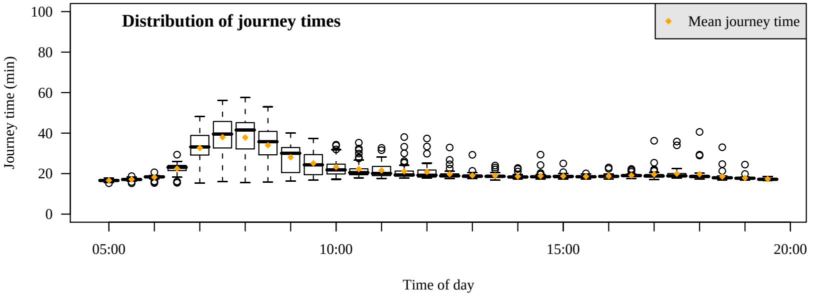
DfT project

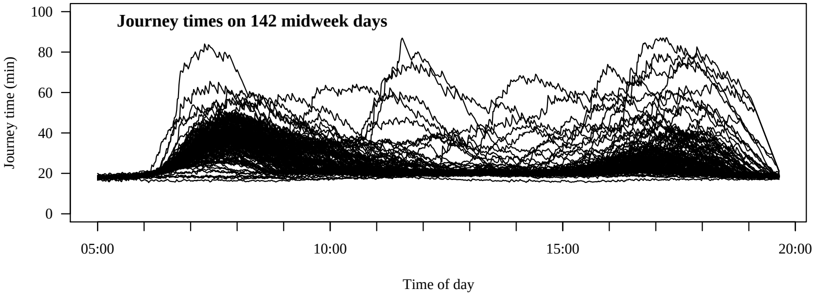
- Now turn to the (short-term) prediction of journey times.
- Report on joint work with Yunus Saatci in a one year project funded by Department for Transport.
- This project emerged from a pilot study by MPhil student Wiebke Werft in 2004/5.
- Focus was to investigate prediction methodolgies (as used in US) with UK MIDAS data.

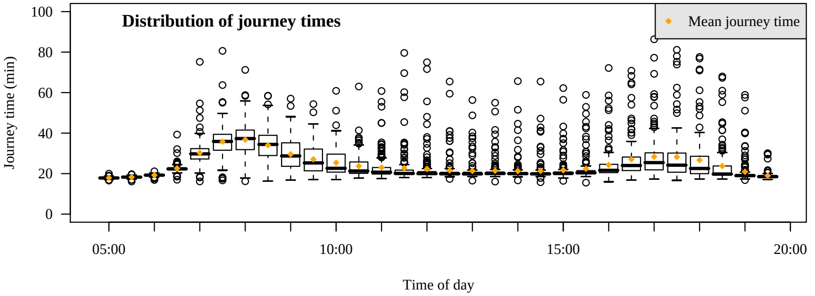
Journey time prediction

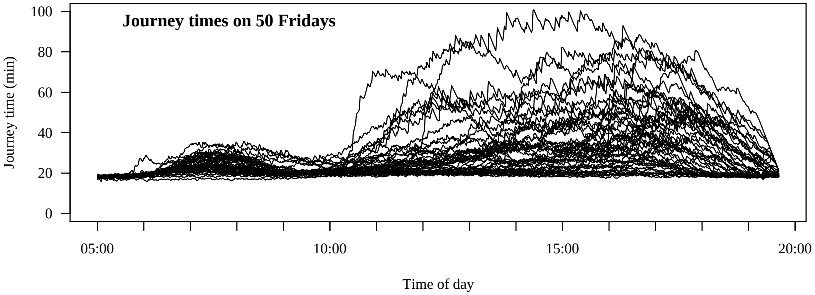


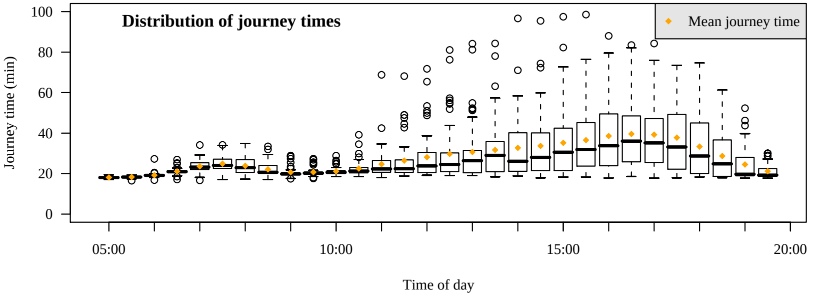

Journey times on 39 Mondays
Journey time (min)
100
80
60
40
20
0
05:00
10:00
15:00
20:00
Time of day


Distribution of journey times
Mean journey time
Journey time (min)
100
80
60
40
20
0
05:00
10:00
15:00
20:00
Time of day

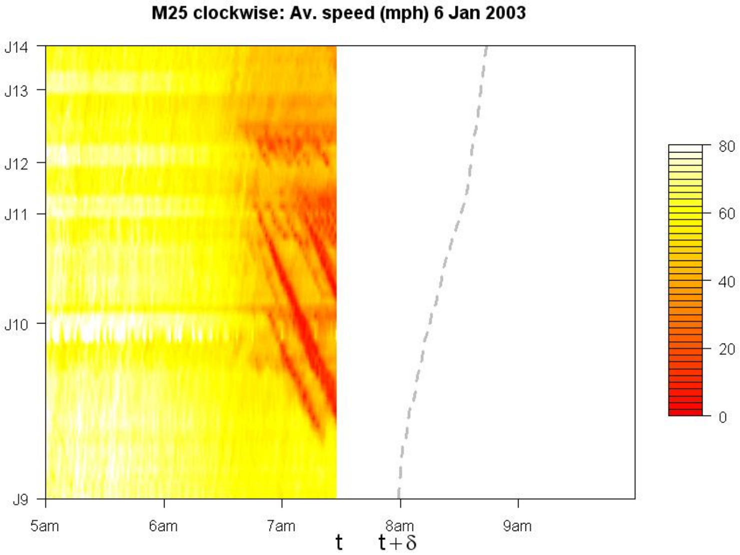

Journey times on 142 midweek days
Journey time (min)
100
80
60
40
20
0
05:00
10:00
15:00
20:00
Time of day


Distribution of journey times
Mean journey time
Journey time (min)
100
80
60
40
20
0
05:00
10:00
15:00
20:00
Time of day

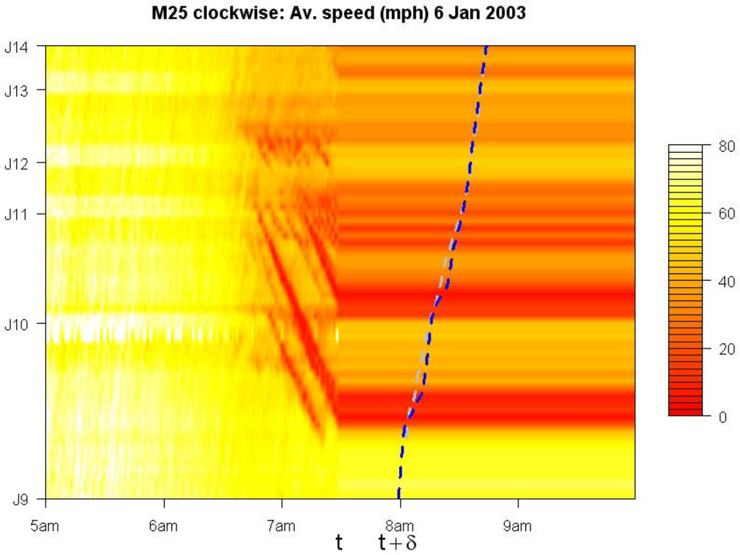

Journey times on 50 Fridays
Journey time (min)
100
80
60
40
20
0
05:00
10:00
15:00
20:00
Time of day


Distribution of journey times
Mean journey time
Journey time (min)
100
80
60
40
20
0
05:00
10:00
15:00
20:00
Time of day

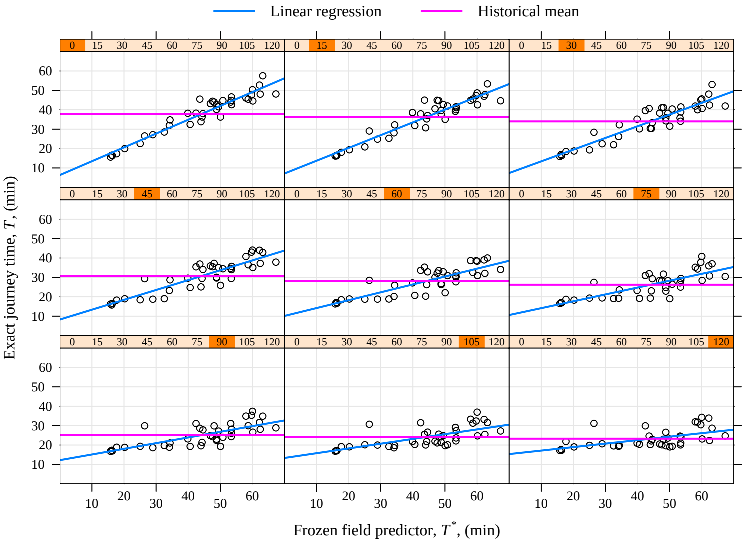
Journey time prediction



Real-time measurements



Linear regression model with varying lags, δ



Some notation and definitions: $T_d(t)$, $T_d^*(t)$, $\overline{T}(t)$

Let $T_d(t)$ be the journey time starting at time t on day $d \in D$
Let the speeds measured at loops/sites $\ell \in \{1, \ldots, L\}$ be $S_d(\ell, t)$ and let the distance between consecutive loops be r.
The frozen field travel time, $T_d^*(t)$, is given by

$$T_d^*(t) = \sum_{\ell=1}^{L-1} \frac{2r}{S_d(\ell, t) + S_d(\ell + 1, t)} \, .$$

The historical mean travel time, $\overline{T}(t)$, for a journey starting at time of day t is given by

$$\overline{T}(t) = \frac{1}{|D|} \sum_{d \in D} T(d, t) \, .$$

Varying coefficients model

Rice and van Zwet (2004) studied a varying coefficients regression model of the form

$$T_d(t+\delta) = \alpha(t,\delta) + \beta(t,\delta)T_d^*(t) + \epsilon$$

where ϵ is a zero mean random variable modelling the random fluctuations and measurement errors.

John Rice and Erik van Zwet (2004) A simple and effective method for predicting travel times on freeways. IEEE Trans on Intelligent Transportation Systems 5(3), 200–207.

Model fitting with smoothed parameters

Smoothed parameters, $(\widehat{\alpha}(t,\delta), \widehat{\beta}(t,\delta))$, may be obtained through a weighted linear regression so as to minimize

$$\sum_{d \in D, s} \left(T_d(s) - \alpha(t,\delta) - \beta(t,\delta) T_d^*(t) \right)^2 K(t + \delta - s)$$

where $K(\cdot)$ denotes a Gaussian density with mean zero and some specified variance σ^2 and s is a general time of day value.

A further predictor

k-Nearest Neighbour

This predictor for $T_d(t+\delta)$ is given in terms of the k closest (past) days $d_1, d_2, \ldots, d_k$ to d in the sense of the distance metric (other metrics are also plausible)

$$m(d, d') = \sqrt{\sum_{t-w \leq s \leq t} \left[T^*_d(s) - T^*_{d'}(s)\right]^2}.$$

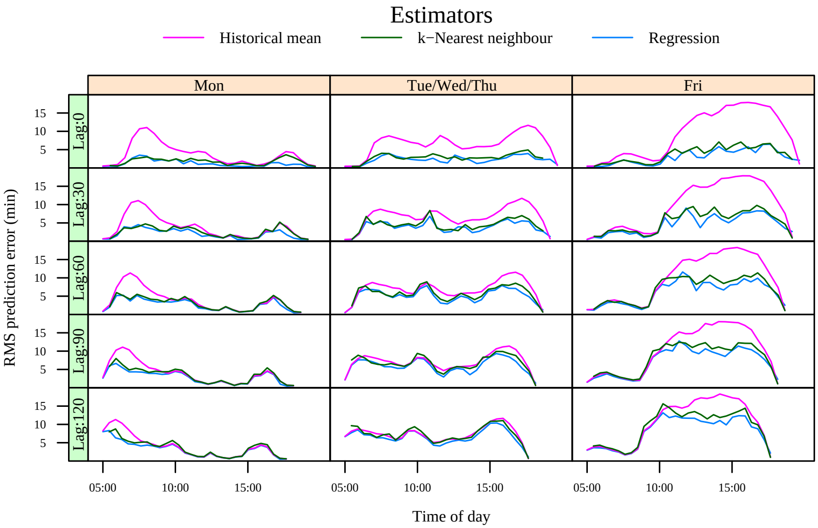
The predictor for $T_d(t+\delta)$ is then

$$T_d^{kNN}(t+\delta) = \sum_{i=1}^{k} w_i T_{d_i}(t+\delta)$$

with weights w_i inversely proportional to distances.
The parameter k and the windowing parameter w help tradeoff the accuracy with the computational overhead.

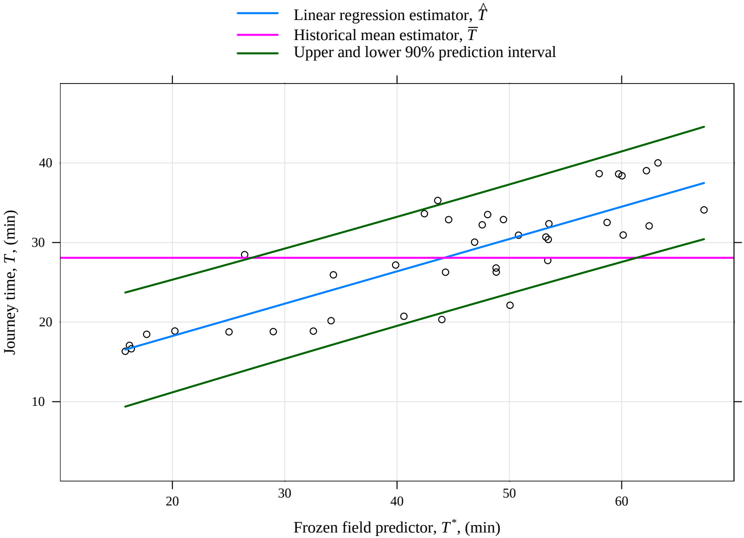
RMS prediction errors

Simple leave-one-out approach



Linear regression model

Prediction intervals, given Normality assumptions (Mondays only for $t = $ 8am and $\delta = $ 60mins)



Discussion

References:

Tom Bellemans (2003) Traffic control on motorways. PhD thesis, Katholieke Universiteit Leuven.

John Rice and Erik van Zwet (2004) A simple and effective method for predicting travel times on freeways. IEEE Trans on Intelligent Transportation Systems 5(3), 200–207.

R.J. Gibbens and W. Werft (2005) Data gold mining: MIDAS and journey time predictors. Significance, 2(3):102–105, September.

R.J. Gibbens and Y. Saatci (2008) Data, modelling and inference in road traffic networks. Proc. R. Soc. Lond. A. Forthcoming. (See: `http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-676.html`)